AD_____

Award Number: W81XWH-12-1-0480

TITLE: Molecular Characterization of Indolent Prostate Cancer

PRINCIPAL INVESTIGATOR: Jun Luo, Ph.D.

CONTRACTING ORGANIZATION: Johns Hopkins University, Baltimore, MD 21218-2680

REPORT DATE: October 2014

TYPE OF REPORT: Annual report

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2014	Annual Report	15Sep 2013 - 14 Sep 2014

4. TITLE AND SUBTITLE
Molecular Characterization of Indolent Prostate Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-12-1-0480

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Jun Luo, Ph.D.

E-Mail: jluo1@jhmi.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Johns Hopkins University,
Baltimore, MD 21218-2680

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Indolent prostate cancers that pose very low risk to aged men occur frequently and may be detected at biopsy, leading to the contemporary problem of prostate cancer over-diagnosis and over-treatment. The objective of the project is to define and characterize indolent prostate cancer using genomic approaches in the clinically relevant context of a cohort meeting the entry criteria for active surveillance. During this funding period, we have determined the optimal technical approaches to RNA sequencing in the specific setting of specimens derived from high-risk and low-risk prostate cancer, collaborated with urologists and pathologists involved in the active surveillance program to define a set of candidate genes to be examined in the active surveillance cohort, and conducted analysis of the longitudinal samples in the tissue repository suitable for the proposed studies. These progresses positioned us for studies designed to develop tools useful in guiding active surveillance.

15. SUBJECT TERMS : nothing listed

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	12	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

Introduction

Indolent prostate cancers that pose very low risk to aged men occur frequently and may be detected at biopsy, leading to the contemporary problem of prostate cancer over-diagnosis and over-treatment. Since progressive acquisition of genomic alterations, both genetic and epigenetic, is a defining feature of all human cancers at different stages of disease progression, RNA and DNA alterations characteristic of indolent prostate tumors may be different from those in clinically significant prostate cancer. However, due to a number of technical constraints, analysis of small volume, very low risk, indolent prostate tumors has not been systemically performed using genome-wide approaches. The primary purpose of the project is to characterize indolent prostate cancer using genomic approaches in the context of a cohort of men predicted to harbor very low-risk prostate cancer at the time of biopsy detection and thus meeting the entry criteria for active surveillance. The scope of the proposed research is: 1) to define the expression signature of indolent prostate cancer by genome-wide expression analysis comparing tissue lesions from very low risk prostate cancer versus high risk prostate cancer defined by pathological outcome measures in men meeting the entry criteria for active surveillance but opting for immediate surgical treatment; 2) to develop a refined signature using biopsy specimens from an active surveillance cohort; and 3) to differentiate indolent prostate cancer from clinically significant prostate cancer using advanced deep-sequencing technologies for both DNA copy number of methylation analysis.

Body

Findings resulting from Task 1: To define indolent human prostate cancer by genome-wide expression analysis comparing tissue lesions from RRP-confirmed very low-risk prostate cancer versus higher-risk prostate cancer (Months 1-24).

Summary: During year 2 of the project period, we focused on technical evaluation of genome-wide approaches utilized for comparison of low-risk and high-risk prostate cancer tissues collected in the standard clinical setting involving formalin-fixation and paraffin embedding (FFPE) of the specimens. We completed two critical project milestones associated with Task 1. First we evaluated the feasibility of using RNA sequencing for genome-wide expression analysis in such specimens and concluded it is feasible to employ this technology which has advantages over traditional microarray-based approaches. On the basis of the findings and the technical trend that was not foreseen at the time of our original grant application, we have slightly revised our approach (see details below). Second, through consultation with our research team we have prioritized a list of candidate genes for inclusion in part of the validation studies in Aim 2 (see details below).

1. RNA-Seq approach for the comparison of low-risk and high-risk prostate tumors. During the project period, the general research field of genome profiling underwent drastic changes. Specifically, RNA sequencing is replacing the traditional expression microarray as the standard methodology for analysis of the entire transcriptome. It is important to adapt to this technical trend. Nevertheless RNA-Seq in paraffin-embedded specimens needs to be fully evaluated under laboratory-specific conditions with full implementation of quality control measures to ensure data validity. We note that additional technical advances have been made that are relevant to RNA-Seq using limited amount of FFPE materials. In studies comparing different RNA-Seq library preparation methods using degraded and/or low-input RNA samples (1, 2), a number of key RNA-Seq technical metrics were evaluated, demonstrating the overall feasibility of achieving 1) efficient rRNA depletion (down to 0.1% of reads aligned to rRNA genes) (1, 2), an essential step in RNA-Seq of FFPE RNA; 2) genome alignment of reads at levels equivalent to RNA-Seq reads from gold-standard high-quality mRNA from fresh frozen samples (1, 2); 3) High sensitivity in transcript detection (1, 2); 4) Acceptable % of exon coverage (greater than 40% of reads mapping to exons) (1, 2); 5) Uniform transcript coverage (1, 2); 6) High concordance in transcript quantification between FFPE RNA-Seq and expression microarrays of fresh frozen tissues(1, 2), at a level similar to the comparison between different expression microarray platforms.

   It is in the context of these latest technical advances that we performed studies evaluating RNA-Seq using FFPE specimens that are used in the comparison of low-risk and high-risk prostate cancer. First we extracted high quality RNA from FFPE specimens (detailed data provided in our previous Progress Report) from two cases (59642 and 59643). In this report, we present summary data derived from these samples. We used 3 different starting amounts (200pg, 2ng 10ng rRNA depleted RNA) of FFPE RNA to make sequencing libraries. We used the rRNA-depletion protocolwith Clontech RobiGone-Mammalian kit(cat#634846 Clontech , USA). After rRNA depletion, cDNA synthesis was made with SMARTer Universal Low Input RNA kit from Clontech. This kit starts with low amount of input RNA then a modified N6 primer (the SMART N6 CDS primer) for first-strand synthesis. The SMARTScribe Reverse Tanscriptase enables template switching and extension to produce the complementary DNA strand. After cDNA amplification, final amplified cDNA is digested with Rsal to remove the SAMRT adapter. Following the Low Input Library Prep kit, FFPE RNA-Seq library was generated. We quantified final libraries with Agilent bioanalyzer and measured with Invitrogen Qubit. All 6 RNA samples were added different indexes to be pooled together for one lane of 50bp single read sequencing. After demultiplexing process with CASAVA, following Clontech

recommendation, additional 7bp sequencing reads (part of SMART adapter) in the beginning of reads were trimmed prior to mapping.

As shown in Table I, Two samples (59642 and 59643) were prepared for sequencing libraries at different starting amounts. All samples were sequenced at about 10 million reads per samples, with mappable read rate around 74-82%, an acceptable measure in most of RNA-seq studies utilizing FFPE specimens. Of note, sequence read duplication rate decreases when starting material amount is lower (from about 74% to 34%), indicating the reduced RNA diversity at lower starting RNA amount. These findings provide important guidance to ongoing studies toward the overall objective of this project. Specifically, the finding suggest that an input amount of 10ng would be desired in ensuing experiments.

Table I: Summary of RNA-Seq mapping results.

RNA Samples	Sample Name	cDNA synthesis starting amount	Total read (millions)	Mappable reads (percent) Millions (%)	Duplication rates (%)
59462	59462-200	200pg	10.05	7.51(74.7%)	73.66
	59462-2	2ng	11.00	8.60(78.2%)	45.59
	59462-10	10ng	10.86	8.19(75.4%)	34.13
59463	59463-200	200pg	10.22	7.81(76.4%)	74.82
	59463-2	2ng	9.67	8.00(82.7%)	54.15
	59463-10	10ng	11.17	9.17(82.1%)	24.5

Next, we measured gene expression levels using TopHat aligner (version 2.0.8) and HTSeq (version 0.5.4). Sequence read counts were then converted to RPKM by considering transcript length and library size. Genes are considered as expressed genes if their expression level (RPKM) is greater than 1.0. Table I summaries the number of genes detected in these experiments. The results are comparable with published literature suggesting overall good quality of RNA-Seq data when limited amount of FFPE tissues are used.

Table II: Number of genes detected by RNA-Seq.

		# of expressed genes	
RNA Samples	Sample Name	RPKM >1	RPKM > 2
59462	59462-200	8,293	7,375

	59462-2	13,332	12,097
	59462-10	14,699	13,087
59463	59463-200	6,778	5,778
	59463-2	12,056	10,931
	59463-10	14,304	12,691
Merged	59462_all	14,120	12,594
	59463_all	13,446	11,908

A number of key performance characteristics were further evaluated to support the feasibility of using FFPE tissues for RNA-Seq for the specific purpose of comparing low-risk and high-risk prostate cancer. Figure 1 shows the mapping rates for exon, intron, and inter-genic sequences. The data suggest minimal effect of the starting amount of RNA on mapping results. Figure 2 shows the % coverage rate for the 5' and 3' of the genes, supporting uniform coverage. Another important measure is % rRNA depletion. Relevant findings on rRNA depletion as a result of input FFPE RNA amount is shown in Figure 3. Sample number 59463 had better rRNA depletion profile than sample number 59462, possibly reflecting better RNA quality (not shown) in 59462. Figure 4 presents Pearson correlation of top 1000 high expression genes between the two low-input samples and sample with 10ng input. The data suggest low data quality in samples with low RNA input. In Figure 5, we present data on average coverage by gene position for the top 1000 expressed genes. Overall, these standard data quality measures support the conclusion that high quality RNA-Seq can be obtained from FFPE RNA in the nanogram range, on the basis of comparable performance characteristics established in current literature.

2. Candidate markers to be tested in Aim 2. A number of markers, including PTEN, ERG, MYC, and ki67, are currently being proposed for testing in Tissue Microarrays before being qualified for expanded studies in the longitudinal active surveillance cohort. We proposes these candidate markers through extensive collaborative consultation, taking into consideration of the literature and studies published by other investigators in the last few years. Candidate markers will be evaluated in Aim 2 using optimized assays (RISH, IHC). We will report the full findings in our final report.

Findings resulting from Task 2: To validate a refined set of genes predictive or indicative of higher-risk disease within a PAS longitudinal cohort (Months 12-36).

According to our project plan in SOW we will carry out studies related to this task during year 2 and year 3 of the project period and will report relevant findings following the studies in our final report. During year 2 of the project period we focus on sample collection. Results of our efforts are summarized below.

We have identified a total of 1060 biopsies suitable for studies proposed in Aim 2. These biopsies met the NCCN very low risk prostate cancer criteria (stage T1c, and PSA <10ng/m; Gleason score <=6; and no more than 2 cores containing cancer, and <=50% of

core involved with cancer; PSA density <0.15ng/ml/g). A subset of them (n=232) represent those from the patients meeting the entry criteria for the active surveillance program but nevertheless reclassified longitudinally.

For the 1060 available research biopsies within our biorepository, diagnostic biopsies, confirmation biopsies and annual monitoring biopsies of follow up patients are all available and factored in the tally. Upon analysis of diagnostic classification distribution of all available biopsies duplicates, there are in total 828 biopsies from 338 cases in the very-low-risk group while there are 232 biopsies from 186 cases in the biopsy progression group. These specimens are more than sufficient for the proposed studies in Aim 2.

Because our slightly revised approach to genome profiling, some of proposed tasks in Aim 2 may have corresponding revision. Specifically we anticipate delay in executing the validation studies, although our candidate marker studies are already underway. Full results and the potential need for a no-cost extension will be communicated before we prepare the Final Report.

Findings resulting from Task 3: To define somatic DNA copy number alterations and methylation changes when higher-risk disease develops in men undergoing PAS (Months 1-36).

Summary: According to our project plan in SOW we initially focused on technical optimization and evaluation of the deep sequencing technology, and will carry out DNA copy number and methylation changes in target specimens from men qualified for active surveillance but that opted for surgery during year 2 of the project period. We have presented our progress on DNA sequencing in our year-1 progress report. Studies on DNA copy number and methylation changes are still ongoing and behind schedule. We will report the findings in our Final Report.

Key Research Accomplishments

1. Established that high quality RNA sequencing data can be generated from limited amount of input RNA isolated from FFPE specimens, for the specific comparison of low-risk and high-risk prostate cancer.
2. Identified candidate markers to be tested in Aim 2.
3. Identified sufficient number of biopsy cases and sections for Aim 2.

Reportable Outcomes

Manuscripts: None at this time.

Presentations: None at this time.

Grant Applications:

Title: Reducing Prostate Cancer Overdiagnosis and Overtreatment (NIH P01, PI: Pienta)
Supporting Agency: NIH/NCI
Performance Period: 7/1/2015 - 6/30/2020
Level of Funding: $310,000
Role: Project Lead, Project 2 (resubmission)
Status: Submitted on Sept. 25th, 2014

Conclusion

High quality RNA sequencing data can be generated from specimens derived from the standard clinical setting for management of patients with low-risk prostate cancer. Foreseeable technical barriers presented by RNA sequencing using low-input and degraded RNA samples have been overcome.

References

1. Adiconis X, Borges-Rivera D, Satija R, DeLuca DS, Busby MA, Berlin AM, Sivachenko A, Thompson DA, Wysoker A, Fennell T, Gnirke A, Pochet N, Regev A, Levin JZ. Comparative analysis of RNA sequencing methods for degraded or low-input samples. Nature methods. 2013;10(7):623-9. Epub 2013/05/21. doi: 10.1038/nmeth.2483. PubMed PMID: 23685885; PubMed Central PMCID: PMC3821180.
2. Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM. Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. BMC genomics. 2014;15:419. Epub 2014/06/04. doi: 10.1186/1471-2164-15-419. PubMed PMID: 24888378; PubMed Central PMCID: PMC4070569.

Appendices

None

Supporting Data (5 figures and figure legends)

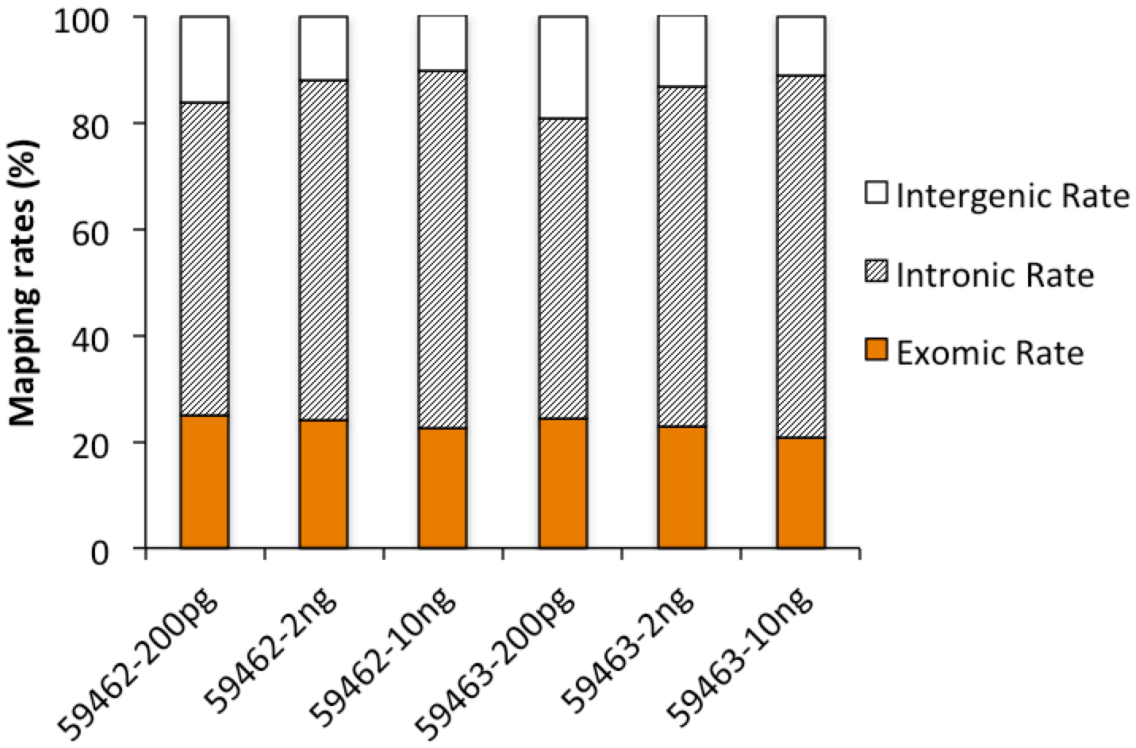

Figure 1: Percentage of sequencing reads mapped to exons, introns, and intergenic regions of the human genome by varying amounts of input FFPE RNA.

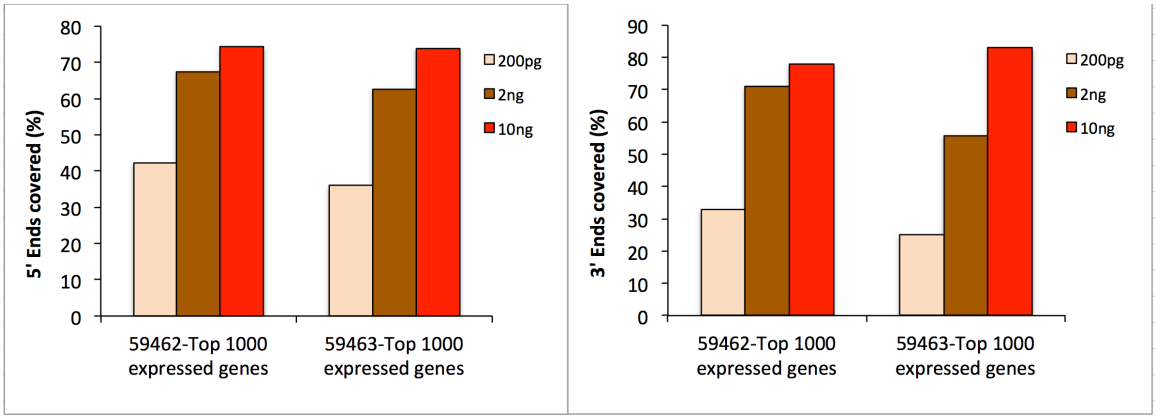

Figure 2: Sequence coverage at the 5' and 3' of the gene transcripts for the top 1000 expressed genes determined by RNA-Seq.

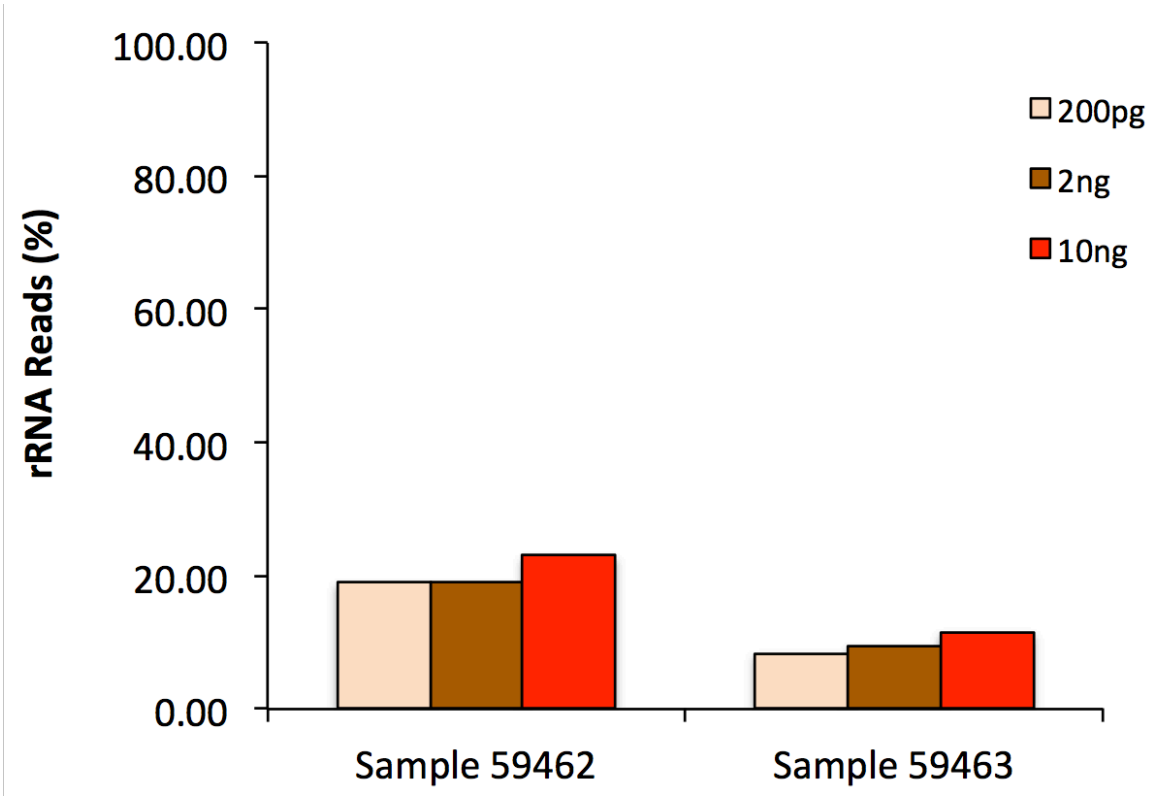

Figure 3. Efficiency of rRNA depletion by sample type and varying amounts of input FFPE RNA.

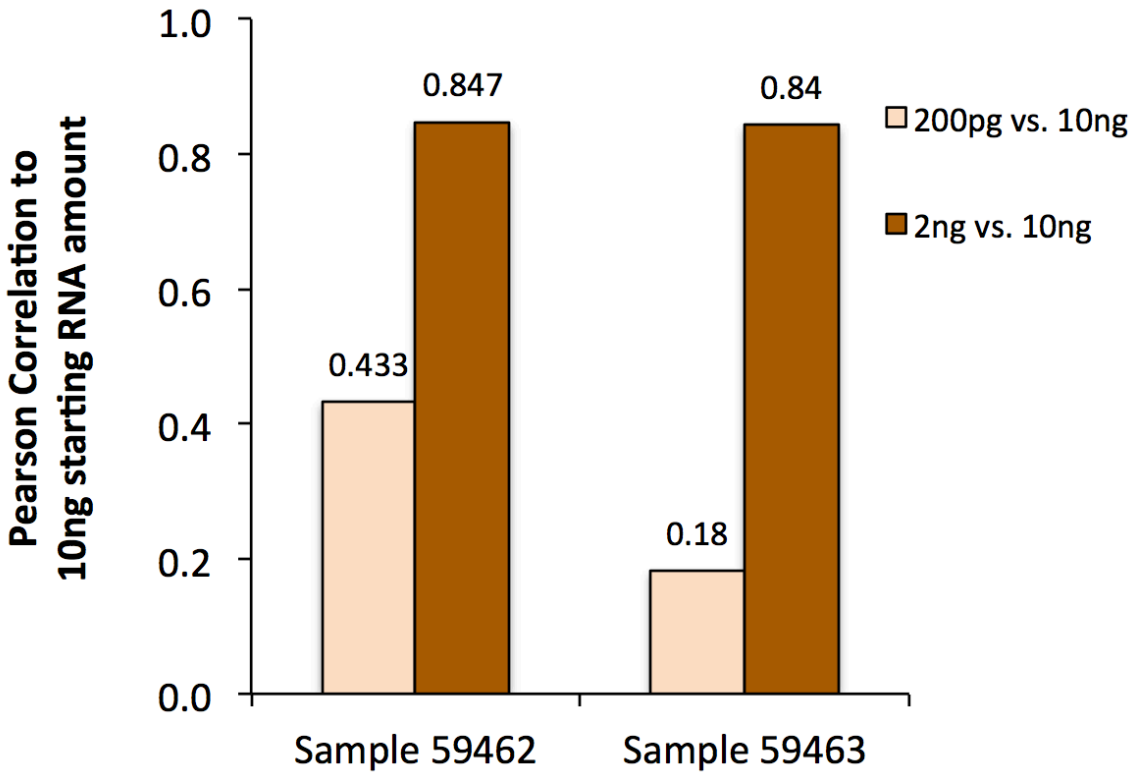

Figure 4: Correlation of transcript abundance between RNA-Seq data derived from lower input RNA (200pg and 2 ng) versus 10ng RNA.

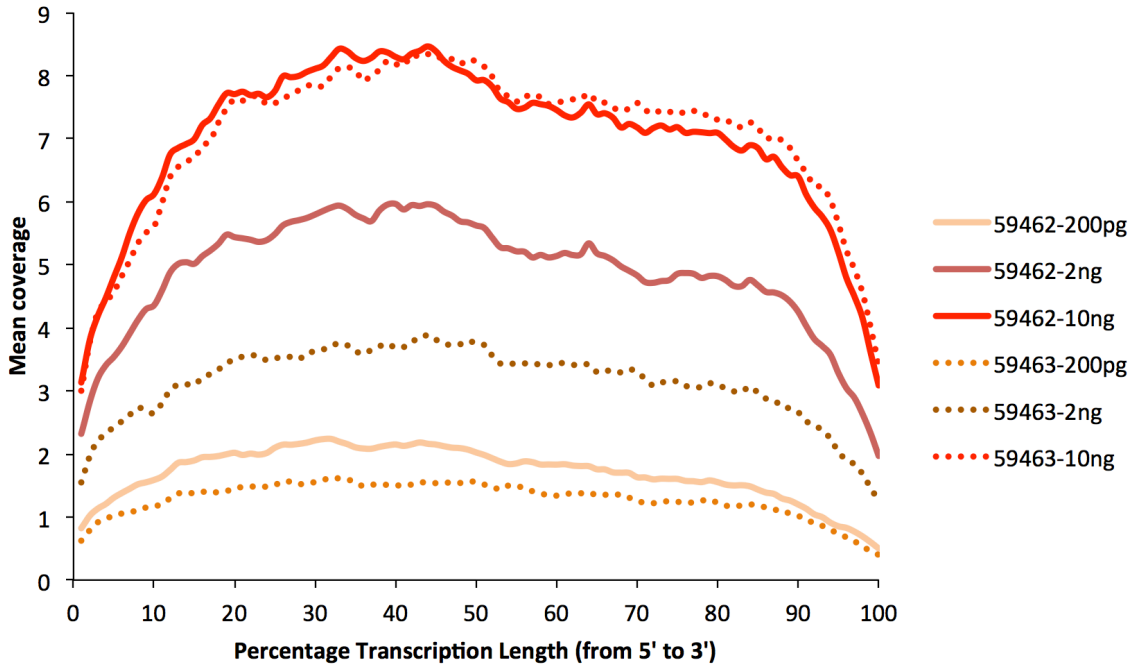

Figure 5: Mean coverage plot by position for top 1000 highly expressed genes determined by RNA-Seq.